Crypt Discovery
Joint Collaboration Activity
20 January 2011

Activity Owners

[redacted] NSA
[redacted] GCHQ

Overview

Our Internet Exploitation capability is built upon our ability to effectively conduct target development and target discovery from IP data. One target discovery strategy is to spot patterns of Internet activity (communications, web browsing, web searching, etc) from which analysts infer suspicious behaviour or intent. Developing Target Detection Identifiers (TDIs), Internet presence data and other key metadata elements are critical enablers for achieving success.

The spread of encryption, particularly Transport Layer Security (TLS), which was previously known as Security Sockets Layer (SSL), and Internet Protocol Security (IPSec), threatens our ability to do effective target discovery/development because pertinent metadata events will be locked within the encrypted channels and difficult, if not impossible, to prise out.

These activities are designed to get a sound understanding of the threat that encryption brings to our ability to do target discovery/development as well as devising mitigations that will (hopefully) allow our Internet Exploitation strategy to prevail.

Plan

Performing discovery within and beyond encryption will require new solutions. We will be investigating and creating solutions in these areas:

1. Monitor the prevalence of encrypted usage within encryption technologies (https, vpn, etc) and across the Internet and into our SIGINT targets' domain. The questions we would like to get into a position to answer are:
    a. What percentage of encryption are we seeing in all traffic? What is the trend over time?
    b. What percentage of encrypted traffic is associated with a target? Trends?
    c. What percentage of targets are using TLS, IPSec, or other encryption technologies? Trends?
    d. How are these figures above broken out across the encrypted web services available (webmail, web searches, geo mapping, web fora, IM, social networking, online shopping, online banking, etc)? Split out the webmail stats per provider. Trends?
    e. What are the most common https sites visited (in all traffic)? Trends?
    f. What are the most common https sites visited by targets? What is the profile of usage? Trends?

Derived From: NSA/CSSM 1-52
Dated: 20070108
Declassify On: 20360101

2. Assess the threat of https to our current exploitation capabilities and to our future intent
    a. Which of our presence and communications metadata (including TDIs) will be lost under encrypted channels?
    b. What is the impact of this loss to our Target Discovery and Target Development tradecraft?
    c. What is the impact of this loss to our ability to cross-correlate TDIs and run more complex queries against our data?

3. Understand and improve our pairing rates within an access and between access points?
    a. What can be paired from what can be viewed?
    b. What could be paired between sites (demonstrate correlation between collaboration environment and TICKETWINDOW sites for example)?
    c. What are the volumes of paired encrypted links and would our CA services be able to handle these?
    d. Can we optimise our passive collection by analysing pairing rates per country and per service level?
    e. How can the cloud enable better pairing by leveraging it's ability to look over time?

4. Research new methods for maintaining an effective TD tradecraft and mitigating the threat of encrypted http traffic.
    a. Are there other TDIs or "TDI-like" identifiers we can develop that can identify users or machines of interest, despite our targets using https? How effective are these? (e.g. are the TDIs persistent across sessions, do they uniquely identify a user or machine, do we always appear in user sessions?)
    b. Can we characterise IPSec sessions from ESP metadata analysis (e.g. using duration, packet length, burstiness)?
    c. How can we best analyze decrypted VPN traffic to develop effective TD tradecraft for VPNs.
    d. Can we use Internet presence or communications metadata with network analysis techniques to identify high-value sessions? (gah: do you think this sits better under (3) above?)
    e. Can we combine knowledge from session processing and content with metadata to enable better methods.

Acquire an understanding for how much more challenging (or easier!) this problem is in the Mobile Internet context, and whether any of the mitigations from (4) above lend themselves to discovering targets behind Mobile Gateways.

Technical Requirements

The ability to understand and track encryption as well as develop new methods as outlined above requires the use of existing metadata from NSA and GCHQ processing system, the ability to create new metadata to test the viability of it's use for the creation of new methods and to enhance existing data, the ability to correlate across what is known (strongly selected content) with metadata and potentially the ability to examine deeper into sessions for evaluation of exploitability. To accomplish the goals above, the following will be used in the JC environment:

- A Joint Collaboration (JC) cloud that contains the following metadata sources:
  - MUTANT BROTH
  - BEARDED PIGGY
  - KARMA POLICE
  - MEMORY HOLE
  - MARBLED GECKO
  - SOCIAL ANIMAL
  - ASDF metadata
- Selected content
  - Related to encryption or specific target sets
- Ability to create new metadata or selected content as needed through XKS or other local processing sources.
- Inclusion of enrichment data
  - TBD
- Ability to interact with CA server internal to the JC environment